# WEB APPLICATION FIREWALL USING DEEP LEARNING ALGORITHM

**MOHD AMIR FARIS BIN MOHD NAWI**



**UNIVERSITI TEKNIKAL MALAYSIA MELAKA**

**BORANG PENGESAHAN STATUS LAPORAN**

JUDUL: WEB APPLICATION FIREWALL USING DEEP LEARNING

SESI PENGAJIAN: [2022 / 2023]

Saya: MOHD AMIR FARIS BIN MOHD NAWI

mengaku membenarkan tesis Projek Sarjana Muda ini disimpan di Perpustakaan Universiti Teknikal Malaysia Melaka dengan syarat-syarat kegunaan seperti berikut:

1. Tesis dan projek adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan unituk tujuan pengajian sahaja.
3. Perpustakaan Fakulti Teknologi Maklumat dan Komunikasi dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. * Sila tandakan (✓)

| | | |
|---|---|---|
| _____ | SULIT | (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972) |
| _____ | TERHAD | (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi / badan di mana penyelidikan dijalankan) |
| ✓ | TIDAK TERHAD | |



(TANDATANGAN PELAJAR)

Alamat tetap: Batalion 17, Pasukan Gerakan AM, Lahad Datu, Sabah

Tarikh: 23/9/2023

(TANDATANGAN PENYELIA)

Dr. Nur Fadzilah binti Othman
Nama Penyelia

Tarikh: 25/9/2023

CATATAN: * Jika tesis ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa.

# WEB APPLICATION FIREWALL USING DEEP LEARNING

MOHD AMIR FARIS BIN MOHD NAWI





This report is submitted in partial fulfillment of the requirements for the Bachelor of Computer Science (Computer Security) with Honours.



FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2023

# DECLARATION

I hereby declare that this project report entitled

**WEB APPLICATION FIREWALL USING DEEP LEARNING**

is written by me and is my own effort and that no part has been plagiarized without citations.

STUDENT : _____ Date : 23/9/2023

(MOHD AMIR FARIS BIN MOHD NAWI)



I hereby declare that I have read this project report and found this project report is sufficient in term of the scope and quality for the award of Bachelor of Computer Science (Computer Security) with Honours.

SUPERVISOR : _____ Date : 25/9/2023

(DR. NUR FADZILAH BINTI OTHMAN)

# DEDICATION

I would want to dedicate my work to my dear parents, whose inspiration and perseverance never stop speaking to me and who consistently give me financial, moral, and spiritual support.

I also dedicate this dissertation to my parents, siblings, grandparents, mentor, and friends who helped me stay inspired to complete this research during the entire process.

Finally, I thank the Great God for helping me finish this project by giving me the strength, courage, and abilities necessary.

## ACKNOWLEDGEMENTS

First and first, I would like to express my sincere gratitude to Universiti Teknikal Malaysia Melaka (UTeM) for giving the undergraduate students in the Faculty of Information and Communication Technology (FICTS) the chance to complete our final year project within the allotted 28 weeks.

My sincere thanks go out to my wonderful supervisor, Dr. Nur Fadzilah binti Othman, for all of her guidance and assurance, which have been essential, and for sparking the bulk of this paper with her perceptive insights.

Above ground, I also want to thank my dear parents, for their encouragement and financial assistance. Last but not least, I want to thank all of my classmates and friends for helping me to finish my research in its entirety.

# ABSTRACT

Web application firewalls (WAFs) are essential for defending web applications against a variety of online dangers. Traditional WAFs, however, frequently find it difficult to keep up with the continually changing threat scenario. Deep learning techniques have been a viable solution to this problem for improving the performance of WAFs. In-depth research on the use of deep learning in web application firewalls is presented in this publication. This research explore the fundamentals of deep learning and its relevance in the context of cybersecurity and WAFs. A detailed analysis of different deep learning architectures, such as Convolutional Neural Networks (CNNs), Long Term Short Memory (LSTM), and support vector machines (SVM), is presented to understand their potential contributions to the WAF's defensive capabilities. Data preprocessing, feature extraction, and model training are all part of the implementation phase for integrating deep learning models into the WAF. Given that web application attack datasets are frequently skewed toward regular traffic, addressing imbalanced datasets is given special consideration. In order to increase model generalization, a number of ways to supplement the sparse labeled data are investigated. The results of this study show that integrating deep learning into WAFs can greatly improve their security capabilities by efficiently identifying and thwarting a variety of web application assaults. The outcomes also emphasize the necessity of ongoing model modification and adaptation to maintain resistance to new dangers. Deep learning is an appealing solution for the changing landscape of web-based cyber threats even if it adds more computational cost and has performance trade-offs.

# ABSTRAK

Tembok api aplikasi web (WAF) adalah penting untuk mempertahankan aplikasi web daripada pelbagai bahaya dalam talian. WAF tradisional, bagaimanapun, sering mendapati sukar untuk mengikuti senario ancaman yang sentiasa berubah. Teknik pembelajaran mendalam telah menjadi penyelesaian yang berdaya maju kepada masalah ini untuk meningkatkan prestasi WAF. Penyelidikan mendalam tentang penggunaan pembelajaran mendalam dalam tembok api aplikasi web dibentangkan dalam penerbitan ini. Laporan itu bermula dengan meneroka asas pembelajaran mendalam dan kaitannya dalam konteks keselamatan siber dan WAF. Analisis terperinci tentang seni bina pembelajaran mendalam yang berbeza, seperti Convolutional Neural Networks (CNN), Long Term Short Memory (LSTM) dan mesin vektor sokongan (SVM), dibentangkan untuk memahami potensi sumbangan mereka kepada keupayaan pertahanan WAF. Prapemprosesan data, pengekstrakan ciri dan latihan model adalah sebahagian daripada fasa pelaksanaan untuk menyepadukan model pembelajaran mendalam ke dalam WAF. Memandangkan set data serangan aplikasi web sering condong ke arah trafik biasa, menangani set data tidak seimbang diberi pertimbangan khusus. Untuk meningkatkan generalisasi model, beberapa cara untuk menambah data berlabel jarang disiasat. Keputusan kajian ini menunjukkan bahawa penyepaduan pembelajaran mendalam ke dalam WAF boleh meningkatkan keupayaan keselamatan mereka dengan cekap dengan mengenal pasti dan menggagalkan pelbagai serangan aplikasi web. Hasilnya juga menekankan keperluan pengubahsuaian dan penyesuaian model yang berterusan untuk mengekalkan ketahanan terhadap bahaya baru. Pembelajaran mendalam ialah penyelesaian yang menarik untuk perubahan landskap ancaman siber berasaskan web walaupun ia menambahkan lebih banyak kos pengiraan dan mempunyai pertukaran prestasi.

# TABLE OF CONTENTS

**PAGE**

# LIST OF TABLES

**PAGE**

# LIST OF FIGURES

**PAGE**

# LIST OF ABBREVIATIONS

| | | |
|---|---|---|
| **FYP** | **-** | **Final Year Project** |
| **WAF** | **-** | **Web Application Firewall** |
| **DL** | **-** | **Deep Learning** |
| **AI** | **-** | **Artificial Intelligence** |
| **SQL** | **-** | **Structured Query Language** |
| **XSS** | **-** | **Cross-Site Scripting** |
| **DDoS** | **-** | **Distributed Denial-of-Service** |
| **IDS** | **-** | **Intrusion Detection System** |
| **HTTP** | **-** | **HyperText Transfer Protocol** |
| **LSTM** | **-** | **Long Short Term Memory** |

# LIST OF ATTACHMENTS

**PAGE**

# CHAPTER 1: INTRODUCTION

## 1.1 Introduction

Web security has become a top priority for businesses as the volume of sensitive data being shared and kept online keeps growing. To identify and stop cyberattacks, traditional web security monitoring techniques like rule-based systems have been widely employed. These systems' weaknesses, however, can result in a high number of false positives and false negatives, which may risk online security. Rule-based web security monitoring relies on pre-determined rules and signatures to detect and prevent attacks. These rules are typically created by security experts and based on known attack patterns. While this approach is effective in identifying and mitigating known attacks, it can be limited in identifying new and unknown attacks, which can lead to false negatives. Additionally, the large number of rules required to cover all potential attack scenarios can lead to a high rate of false positives, which can overwhelm security teams and lead to alert fatigue. The limitations of rule-based web security monitoring create a need for a comprehensive analysis of the effectiveness of this approach in protecting web applications and data from cyber threats. To encounter this problem, Deep Leaning-based web security monitoring can be introduced.



## 1.2 Problem Statement

The limitations of traditional rule-based web security monitoring techniques, such as high false positive and false negative rates, pose significant risks to online security as the volume of sensitive data shared and stored online continues to grow. These systems heavily rely on pre-determined rules and signatures created by security experts, making them effective in detecting known attacks but limited in identifying

new and unknown threats. Furthermore, the large number of rules required to cover all potential attack scenarios often leads to a high rate of false positives, overwhelming security teams and causing alert fatigue. Therefore, there is a critical need to address these limitations and evaluate the effectiveness of rule-based approaches in protecting web applications and data from cyber threats. Introducing Deep Learning-based web security monitoring may offer a potential solution to enhance the detection and prevention of attacks by leveraging advanced machine learning techniques.

## 1.3 Project Question

Project question is usually derived from problem statement. It is important to know the questions to solve the problem statement.

**Table 1.1: Table of Project Question**

| No. | Project Question |
|---|---|
| 1 | What type of Deep learning algorithm used to identify the malicious packet? |
| 2 | How to ensure our system has low error rate? |
| 3 | How can user know the detail of the attack? |



## 1.4 Project Objective

As this project is developed, a number of objectives are being discussed.

**Table 1.2: Table of Project Objective**

| No. | Project Objective |
|---|---|
| 1 | Identify the best Deep Learning algorithm can be used for WAF |
| 2 | Implement the highest accuracy Deep Learning model in detecting malicious payloads. |
| 3 | Develop a WAF with dashboard that able to display the attack details. |

## 1.5 Report Organization

The first chapter demonstrates how deep learning may be used to identify online attacks. This part also contains the issue description, project question, project aim, project scope, and project contribution.

The second chapter discusses several peer-reviewed studies on implementing Deep Learning technique on Web Application Firewall (WAF) in identifying web attack. It covers known WAF categorizing techniques, along with their strengths, weaknesses, and limits, and also proposed strategies and improvements to existing techniques.

The methods and strategies used to complete this project are supported in Chapter 3. To ensure that the given tasks are completed on time, this chapter also includes a research milestone and a Gantt chart.

Chapter 4 will defines the user interface of the project, the requirement and along with the design of the system.

For chapter 5, the programming and creation of deep learning algorithms for web attack detection are the main topics of this chapter. The anticipated outcome of the system will also be covered in this chapter.



For chapter 6, it will describes about the activities involved in testing phase of the software, the project outcomes and performance of the Web Application Firewall system.

Chapter 7 will summarize and wraps up the whole project as well as the strength and limitations involved. Future enhancement and the contribution of the project will also be outlined in this chapter.

## 1.6 Conclusion

In conclusion, the goal of this project is to assist in securing websites from hackers. Additionally, this project can make it easier for users to identify the attack's specifics utilizing the dashboard that was created.

The next chapter's literature study will go through research papers on Web Application Firewalls and their implementation strategies.

# CHAPTER 2: LITERATURE REVIEW AND PROJECT METHODOLOGY

## 2.1 Introduction

This section will assess some relevant studies or articles on the topic of Web Application Firewall Using Deep Learning. This literature review is based on a few sources, including journal articles and conference papers. The methods and procedures used in this study will be described as the research domain. This chapter will also discuss previous or current methods or procedures, as well as their shortcomings or constraints. The comparability of earlier research methodologies or techniques will be covered in this section.

## 2.2 Web Application Firewall(WAF)



A WAF, or web application firewall, assists in securing online applications by screening and keeping track of HTTP traffic between a web application and the Internet. Typically, it defends online applications against threats like SQL injection, file inclusion, cross-site scripting (XSS), and cross-site forgery. A WAF is not made to withstand all kinds of attacks; it is a protocol layer 7 defense (in the OSI model).

A WAF operates by examining HTTP traffic between web clients and web servers, inspecting the content of HTTP requests and responses, and filtering out malicious traffic that could exploit vulnerabilities in the web application. This can be done by applying a set of predefined rules or policies to filter out known attack patterns, or by using machine learning and artificial intelligence techniques to identify and block new and unknown threats.

A WAF can be deployed as a hardware appliance, a virtual appliance, or as a cloud-based service. It can be integrated with other security solutions such as intrusion detection and prevention systems (IDPS) and security information and event management (SIEM) systems to provide a comprehensive security solution for web applications.

WAFs are important for businesses that rely on web applications for critical functions, such as online transactions, customer interactions, and other sensitive operations. By implementing a WAF, organizations can reduce the risk of data breaches, protect their reputation, and comply with regulatory requirements for data security. While a WAF can help to secure web applications, it is important to note that it is not a silver bullet and must be used in conjunction with other security measures such as regular vulnerability scans, penetration testing, and secure coding practices.

### 2.2.1 Cloud -Based WAF

A cloud-based WAF is a quick-deploy, turnkey deployment option that is affordable and simple to implement. Cloud-based WAFs are typically subscription-based and have low up-front costs. Cloud-based WAFs have access to regularly updated threat intelligence and could also provide managed services to assist you in defining security policies and defending against attacks as they happen.



A cloud-based WAF should ideally give users the choice between in-line and out-of-path (OOP) service deployment. It is possible to optimise an API-based OOP deployment for multi-cloud environments, on-premises environments, hybrid environments, etc. by taking advantage of a number of distinctive advantages.

Cloud-based WAFs have gained popularity among businesses of all kinds, from large corporations to tiny firms, because they may offer high levels of security for low upfront expenses and don't require a lot of in-house security experience.

First of all, it is very cost-effective, negating the need for expensive on-premises hardware and maintenance costs. The setup procedure is streamlined by its simplicity in deployment and implementation, which requires little initial outlay. Additionally, it offers constant levels of security, as well as centralized management

and reporting capabilities, assuring consistent security measures across different environments, making it an effective and affordable option for protecting online applications.

### 2.2.2 Software-Based WAF

An alternative to a hardware-based WAF is a software-based WAF. A WAF that is software-based operates locally (on-premise), in a private cloud, or in a public cloud as a virtual appliance or an agent. Other WAFs are also available that are made specifically to be embedded in environments that use containers to host microservices, like Kubernetes.

Without a hardware device, a software-based WAF is installed on top of a virtual machine (VM). A virtual machine is an environment where multiple users can use a computer system at the same time by dividing the system as if there were several small computer systems.

Particularly when a business has in-house security experience and resources, a software-based Web Application Firewall (WAF) offers distinct advantages, including more customization choices. Additionally, compared to hardware-based WAF solutions, it offers cheaper initial, deployment, and continuing maintenance costs, making it a cost-effective option for customizing web application security to particular demands while easing financial concerns.



### 2.2.3 Hardware-Based WAF

This type of WAF locally installed on a network is a hardware-based WAF, also referred to as a network-based WAF. Due to the need for maintenance and storage space, these WAFs are typically the most expensive. Their main goal is to reduce latency.

Hardware-based WAF have become less and less relevant in recent years as cloud-based WAF have largely replaced them.